INSTALLING AND RUNNING BOWTIE, TOPHAT, AND CUFFLINKS

These instructions are written for Mac OS X. Some modifications may be needed for other operating systems. Text in italics is user-specific. Orange and green colored text are used to delineate text separated by a space.

Note: steps for running the software are in their simplest forms.

1. **Download Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml):**
   Decompress archived files and move the bowtie folder into the directory you would like to store it in (e.g. create a new directory called Bowtie and move the folder into that directory). Installation is complete.

2. **Download TopHat (http://ccb.jhu.edu/software/tophat/tutorial.shtml):**
   Decompress archived files and move the bowtie folder into the directory you would like to store it in (e.g. create a new directory called TopHat and move the folder into that directory). Installation is complete.

3. **Download Cufflinks (http://cufflinks.cbcb.umd.edu/tutorial.html):**
   Decompress archived files and move the Cufflinks folder into the directory you would like to store it in (e.g. create a new directory called Cufflinks and move the folder into that directory). Installation is complete.

4. **Set the home environment to point to bowtie2, TopHat, and Cufflinks so that the software can be run from anywhere on your computer (you may need to edit path by placing a colon before /bin):**

   Bowtie2
   ```
   bash$ echo "BT2_HOME=path_to_bowtie2_directory" >>~/.bash_profile
   bash$ echo "export BT2_HOME" >>~/.bash_profile
   bash$ echo "export PATH=$PATH:$BT2_HOME/bin" >>~/.bash_profile
   bash$ source ~/.bash_profile
   ```

   TopHat
   ```
   bash$ echo "TH_HOME=path_to_tophat_directory" >>~/.bash_profile
   bash$ echo "export TH_HOME" >>~/.bash_profile
   bash$ echo "export PATH=$PATH:$TH_HOME/bin" >>~/.bash_profile
   bash$ source ~/.bash_profile
   ```

   Cufflinks
   ```
   bash$ echo "CL_HOME=path_to_cufflinks_directory" >>~/.bash_profile
   bash$ echo "export CL_HOME" >>~/.bash_profile
   bash$ echo "export PATH=$PATH:$CL_HOME/bin" >>~/.bash_profile
   bash$ source ~/.bash_profile
   ```

5. **Download a prebuilt genome index from Illumina (http://support.illumina.com/sequencing/sequencing_software/igenome.html) or build your own:**

   Obtain the fasta formatted genome sequence for your species of interest and then run the following command from the directory that contains the genome sequence. Note: genome.fa is your fasta formatted genome sequence. A genome can be any collection of sequences (e.g. if your species doesn't have a complete genome, a collection of contigs can be used).

   ```
   bash$ bowtie2-build genome.fa genome_name
   ```

6. **If necessary, remove adapter sequences:**
   Adapter removal can be done using the fastx_clipper program in the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) or Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic).

7. **If necessary, trim reads to remove poor quality bases:**
   Sequences can be trimmed using the fastq_quality_trimmer program in the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) or Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic).

8. **Place a fasta formatted genome sequence for your species of interest in the bowtie folder that contains your indexed genome. This is typically the fasta file you used to build a bowtie index or can be obtained from Illumina along with the indexed genome (see #5 above).**

9. **Run TopHat to map reads:**

   Run TopHat on each library but can combine multiple fastq files from the same library.

   bash$ cd *directory_containing_fastq_file*
   bash$ tophat -r *mate_distance* -o *output_folder_sampleX*
   path_to_bowtie_index_for_reference_genome/index_prefix fastq_fileX

   TopHat produces several files. The accepted_hit.bam file is used by Cufflinks.
   For commonly used options, see the TopHat manual (bash$ tophat).

10. **Optional: run Cufflinks to assemble transcripts** (cufflinks uses the accepted_hits.bam output files from TopHat)**:**

    Run Cufflinks on each library from the TopHat alignment.

    bash$ cufflinks -o *cufflinks_output_sampleX path_to_libraryX_accepted_hits.bam*

11. **Optional (include if step 10 is included): run Cuffmerge to merge the assembled transcripts from replicate libraries**:

    Create a file called assemblies.txt with the paths to each of the individual assemblies files (shown is an example of merging two assemblies):

    bash$ echo *./cufflinks_output_sampleX*/transcripts.gtf >assemblies.txt
    bash$ echo *./ cufflinks_output_sampleX*/transcripts.gtf >>assemblies.txt

    Merge assemblies using Cuffmerge:
    bash$ cuffmerge -s *path_to_genome_fasta_file* assemblies.txt

    You should now have a single file, merged.gtf located in a folder called merged_asm that was created by cuffmerge, that contains all of the predicted transcripts based on the sequencing data.

12. **Run Cuffdiff to identify genes differentially regulated between samples:**

bash$ *cuffdiff file_containing_gene_annotations.gtf library1_replicate1/*
accepted_hits.bam,./library1_replicate2/accepted_hits.bam \library2_replicate1/
accepted_hits.bam,./library2_replicate2/accepted_hits.bam

For *file_containing_gene_annotations.gtf*, if working with a species with an annotated genome, you can use the canonical gene annotations. If working with a non-model species, use the output from assemblies from Cufflinks and Cuffmerge.
The accepted_hits.bam files are the output files from TopHat, step 9.

Several output files are generated. The gene_exp.diff is the summary file and can be opened and viewed in excel
Sort the data in Excel based on the q value.
Identify which genes are significantly different between samples.